SVM Soft Margin Classification

Nipun Batra

July 21, 2025

IIT Gandhinagar

"SLIGHTLY" NON-SEPARABLE DATA
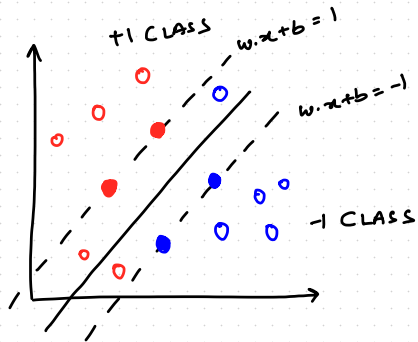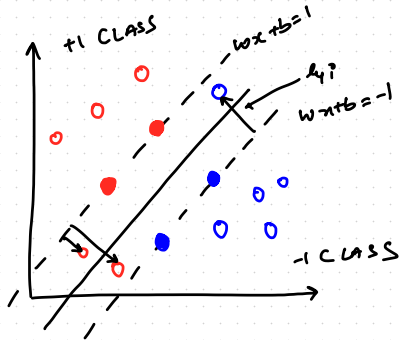
+1 CLASS

$w \cdot x + b = 1$

$w \cdot x + b = -1$

−1 CLASS

- Can we learn SVM for "slightly" non-separable data without projecting to a higher space?
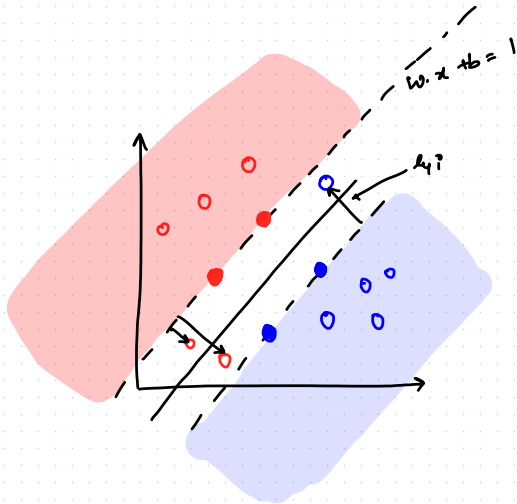
Soft-Margin SVM

- Can we learn SVM for "slightly" non-separable data without projecting to a higher space?

- Introduce some "slack" (ξ_i) or loss or penalty for samples - allow some samples to be misclassified

"SLIGHTLY" NON-SEPARABLE DATA

+1 CLASS

$wx + b = 1$

ℓ_{4i}

$wx + b = -1$

-1 CLASS

ℓ_{4i}: Distance from margin

ZONE 1

$y_i = +1$

$w \cdot x_i + b \geqslant 1$

$w \cdot x + b = 1$

ξ_i

ZONE 1

$w \cdot x + b = 1$

ξ_i

$y_i (\vec{w} \cdot \vec{x_i} + b) > 1$
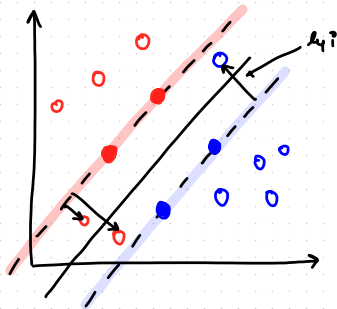
LOSS$_i = 0$ ($\xi_i = 0$)

$\Downarrow$

NO PENALTY
(OR SLACK NEEDED)

ZONE 2

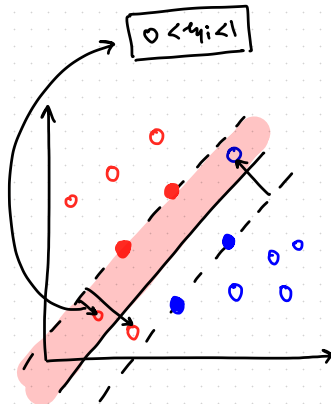$$y_i (\vec{w} \cdot \vec{x}_i + b) = 1$$

$$LOSS_i = 0$$

$$(\xi_i = 0)$$

$$0 < u_i < 1$$

ZONE 3

$y_i (\vec{w} \cdot \vec{x_i} + b) < 1$

$LOSS_i \neq 0$ $(0 < u_i < 1)$

POINT CORRECTLY
CLASSIFIED

(BUT WRONG
SIDE OF MARGIN)

$0 < \xi_i < 1$

$\vec{w} \cdot \vec{x} + b = 1$ $\vec{w} \cdot \vec{x} + b = 1 - \xi_i$

ξ_i
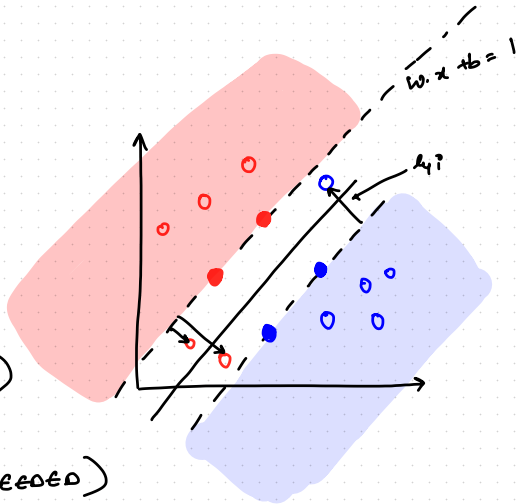
$\vec{w} \cdot \vec{x} + b = 0$

ZONE 3

$y_i (\vec{w} \cdot \vec{x}_i + b) < 1$

$\text{LOSS}_i \neq 0 \quad (0 < \xi_i < 1)$
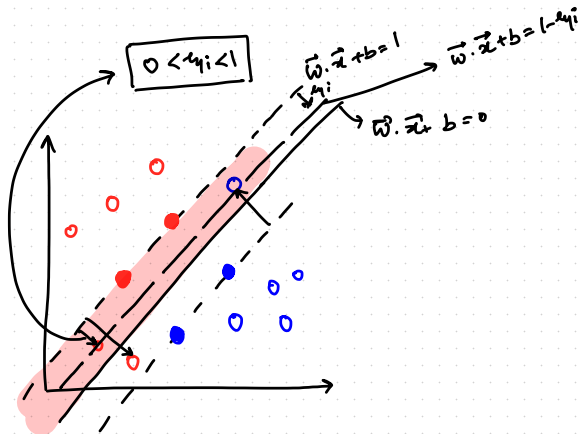
POINT CORRECTLY
CLASSIFIED

(BUT WRONG
SIDE OF MARGIN)

ZONE 4

$$y_i (\vec{w} \cdot \vec{x_i} + b) < 1$$

POINT INCORRECTLY
   CLASSIFIED
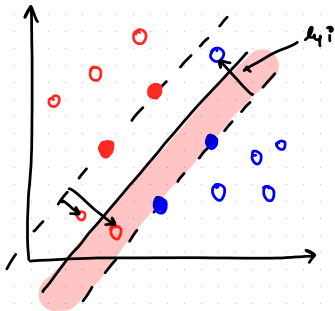
LOSS $i \neq 0$
$\xi_{y_i} > 1$

ZONE 4

$$y_i \left(\vec{w} \cdot \vec{x}_i + b \right) < 1$$

POINT INCORRECTLY CLASSIFIED

LOSS$_i \neq 0$

$\xi_i > 1$
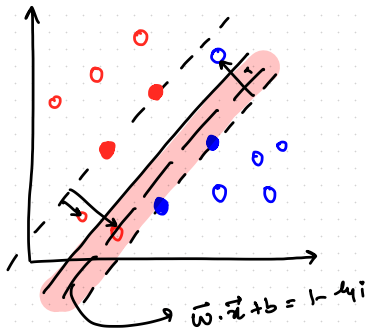
$\vec{w} \cdot \vec{x} + b = 1 - \xi_i$

Change Objective

$\min \frac{1}{2}||\bar{w}||^2 + C \sum_{i=1}^{n} \xi_i$

s.t. $y_i(\bar{w}\bar{x}_i + b) \geq 1 - \xi_i$

Soft-Margin SVM

Change Objective

$\min \frac{1}{2}||\bar{w}||^2 + C \sum_{i=1}^{n} \xi_i$

s.t. $y_i(\bar{w}\bar{x}_i + b) \geq 1 - \xi_i$

In Dual:
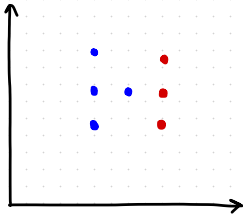
$$\text{Minimize} \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j$$

s.t.

$$0 \leq \alpha_i \leq C \quad \& \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

BIAS- VARIANCE   TRADE-OFF

BIAS- VARIANCE   TRADE-OFF

$\vec{w} \cdot \vec{x} + b = 0$
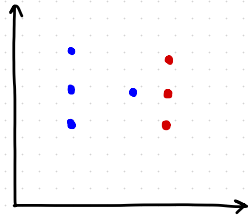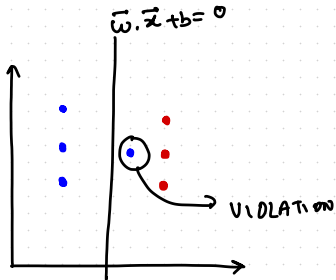
VIOLATION

LOW   C

LOW   PENALTY   FOR   VIOLATION

HIGH   TRAIN   ERROR

HIGH   BIAS

BIAS- VARIANCE   TRADE-OFF

$\vec{w}.\vec{x}+b= 0$
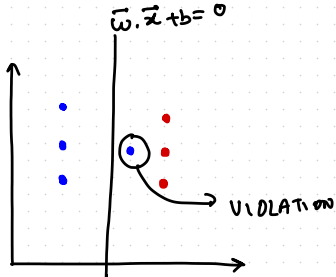
VIOLATION

LOW   C

LOW   PENALTY   FOR   VIOLATION

HIGH   TRAIN   ERROR

HIGH   BIAS

BIG   MARGIN

$\vec{w}.\vec{x}+b= 0$
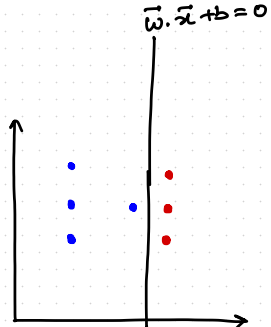
HIGH   C

HIGH   PENALTY   FOR   VIOLATION

HIGH   VARIANCE

SMALL   MARGIN

Bias Variance Trade-off for Soft-Margin SVM

Low C $\implies$ Higher train error (higher bias)

High C $\implies$ Very sensitive to datasete (high variance)

If $C \to 0$

    Objective $\to$ min $\frac{1}{2}||\bar{w}||^2$

    $\implies$ Choose large margin (without worrying for ξ_is)

$$\boxed{\text{Recall: Margin} = \frac{2}{||\bar{w}||}}$$

If $C \to \infty$ (or very large)   Objective $\to$ min $C \sum \xi_i$ or choose W, b, s.t. ξ_i is small!

Soft-Margin SVM

Q) What is the equivalent of hard margin?

  a $C \rightarrow 0$

Soft-Margin SVM

Q) What is the equivalent of hard margin?

a $C \to 0$

b $C \to \infty$

Soft-Margin SVM

Q) What is the equivalent of hard margin?

  a $C \to 0$

Q) What is the equivalent of hard margin?

a $C \to 0$

b $\boxed{C \to \infty}$ $\implies$ No violations!!

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\bar{w}\bar{x}_i + b) = 1$

$\therefore$ As C increases, # support vectors decreases

Notebook: SVM-soft-margin

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\bar{w}\bar{x}_i + b) = 1$
- Zone 3: $0 < \xi_i < 1$ (correctly classified)

$\therefore$ As C increases, # support vectors decreases

Notebook: SVM-soft-margin

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\bar{w}\bar{x}_i + b) = 1$
- Zone 3: $0 < \xi_i < 1$ (correctly classified)
- Zone 4: $\xi_i > 1$ (Misclassified)

$\therefore$ As C increases, $\#$ support vectors decreases

Notebook: SVM-soft-margin

SVM Formulation in the Loss + Penalty Form

Objective:

$$\min \frac{1}{2}||\bar{w}||^2 + C\sum_{i=1}^{N}\xi_i$$

Now:

$$y_i(\bar{w}\bar{x}_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 1 - y_i(\bar{w}\bar{x}_i + b)$$

But $\xi_i \geq 0$

$$\therefore \xi_i = \max\left[0, 1 - y_i(\bar{w}\bar{x}_i + b)\right]$$
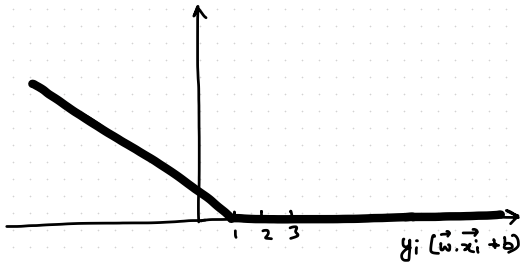
SVM Formulation in the Loss + Penalty Form

$\therefore$ Objective is:
$$\min C \sum \xi_i + \frac{1}{2}||\bar{w}||^2$$

$$\implies \min C \sum_{i=1}^{N} \max \left[0, 1 - y_i(\bar{w}\bar{x}_i + b)\right] + \frac{1}{2}||\bar{w}||^2$$

$$\implies \min \underbrace{\sum_{i=1}^{N} \max \left[0, 1 - y_i(\bar{w}\bar{x}_i + b)\right]}_{\text{Loss}} + \underbrace{\frac{1}{2C}||\bar{w}||^2}_{\text{Regularisation}}$$

HINGE LOSS

$y_i (\vec{w} . \vec{x_i} + b)$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^{N} \max\left[0, 1 - y_i(\bar{w}\bar{x}_i + b)\right]$

- Case I $\quad y_i(\bar{w}\bar{x}_i + b) = 1$
  Lies on Margin: $Loss_i = 0$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^{N} \max \left[0, 1 - y_i(\bar{w}\bar{x}_i + b) \right]$

- Case I $\quad y_i(\bar{w}\bar{x}_i + b) = 1$
  Lies on Margin: $Loss_i = 0$

- Case II
  $$y_i(\bar{w}\bar{x}_i + b) > 1$$
  $Loss_i = 0$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^{N} \max \left[0, 1 - y_i(\bar{w}\bar{x}_i + b)\right]$

- Case I $\quad y_i(\bar{w}\bar{x}_i + b) = 1$
  Lies on Margin: $Loss_i = 0$

- Case II
  $$y_i(\bar{w}\bar{x}_i + b) > 1$$
  $Loss_i = 0$

- Case III
  $$y_i(\bar{w}\bar{x}_i + b) < 1$$
  $Loss_i \neq 0$

Hinge Loss Continued

Q) Is hinge loss convex and differentiable?
  Convex: ✓
  Differentiable: X
  Subgradient: ✓

SVM Loss is Convex

Hinge Loss $\sum(\max[0, (1 - y_i(\bar{w}x_i + b))])$ is convex

Penalty $\frac{1}{2}||\bar{w}||^2$ is convex

$\therefore$ SVM loss is convex